IEEE Xplore RELEASE 2.1

Welcome United States Patent and Trademark Office

☐ Search Session History

BROWSE   **SEARCH**   **IEEE XPLORE GUIDE**   **SUPPORT**

Fri, 13 Jan 2006, 12:39:41 PM EST

Search Query Display

Edit an existing query or compose a new query in the Search Query Display.

Select a search number (#) to:

- Add a query to the Search Query Display
- Combine search queries using AND, OR, or NOT
- Delete a search
- Run a search

	Recent Search Queries	Results
#1	((nearest and neighbor and cluster and document and precision and recall)<in>metadata)	0
#2	((nearest and neighbor and cluster and document and (precision or recall))<in>metadata)	0
#3	((nearest and neighbor and cluster and document) <in>metadata)	13
#4	(nearest <and>neighbor)<and>cluster	8868
#5	((nearest <and>neighbor) <and>cluster<and>precision<and>recall<IN>metadata)	22
#6	(nearest <and>neighbor)<and>cluster	8868
#7	((nearest <and>neighbor)<and>cluster<IN>metadata)	1773
#8	(nearest <and>neighbor)<and>cluster	8868
#9	((nearest <and>neighbor)<and>cluster<IN>metadata)	1773
#10	(nearest <and>neighbor)<and>cluster	8868
#11	((nearest <and>neighbor)<and>cluster<IN>metadata)	1773
#12	(nearest <and>neighbor)<and>cluster	8868
#13	(nearest <and>neighbor)<and>cluster	8868
#14	((nearest <and>neighbor)<and>cluster<IN>metadata)	1773
#15	((nearest <and>neighbor)<and>cluster<IN>metadata)	1773
#16	(nearest <and>neighbor)<and>cluster	8868
#17	((nearest and neighbor and cluster and document and precision and recall)<in>pdfdata)	165
#18	((nearest and neighbor and cluster and document and precision and recall)<in>pdfdata)	165
#19	((nearest and neighbor and cluster and document and precision and recall)<in>pdfdata)	165

((nearest and neighbor and cluster and document and precision

Help   Contact Us   Privacy & Security   IEEE.org

Indexed by
Inspec

Subscribe (Full Service) Register (Limited Service, Free) Login

P⦿RTAL

USPTO

Search: ⦿ The ACM Digital Library   ○ The Guide

+nearest +neighbor +cluster +document +abstract:nearest +

SEARCH

THE ACM DIGITAL LIBRARY

ⁱᶜ Feedback Report a problem Satisfaction survey

Terms used
nearest neighbor cluster document nearest neighbor cluster document precision recall

Found **3 of 552**
searched out of 552.

Sort results by | relevance ▼ |

Display results | expanded form ▼ |

❤ Save results to a Binder

? Search Tips

☐ Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 3 of 3

Relevance scale ☐ ▄ ▄ ■ ■

1   A tree algorithm for nearest neighbor searching in document retrieval systems   ▄

Caroline M. Eastman, Stephen F. Weiss
May 1978 **ACM SIGIR Forum , Proceedings of the 1st annual international ACM SIGIR conference on Information storage and retrieval SIGIR '78**, Volume 13 Issue 1
Publisher: ACM Press

Full text available: 🔁 pdf(651.08 KB)   Additional Information: full citation, abstract, references, citings, index terms

> The problem of finding nearest neighbors to a query in a document collection is a special case of associative retrieval, in which searches are performed using more than one key. A nearest neighbors associative retrieval algorithm, suitable for document retrieval using similarity matching, is described. The basic structure used is a binary tree, at each node a set of keys (concepts) is tested to select the most promising branch. Backtracking to initially rejected branches is allowed and ofte ...

2   Hierarchic document classification using Ward's clustering method   ▄

A. El-Hamdouchi, P. Willett
September 1986 **Proceedings of the 9th annual international ACM SIGIR conference on Research and development in information retrieval**
Publisher: ACM Press
Full text available: 🔁 pdf(974.30 KB)   Additional Information: full citation, abstract, references, citings

> In this paper, we discuss the application of a recent hierarchic clustering algorithm to the automatic classification of files of documents. Whereas most hierarchic clustering algorithms involve the generation and updating of an inter-object dissimilarity matrix, this new algorithm is based upon a series of nearest neighbor searches. Such an approach is appropriate to several clustering methods, including Ward's method which has been shown to perform well in experimental studies of hierarch ...

3   Database techniques for archival of solid models   ▄

David McWherter, Mitchell Peabody, Ali C. Shokoufandeh, William Regli
May 2001 **Proceedings of the sixth ACM symposium on Solid modeling and applications**
Publisher: ACM Press

Full text available: 🔁 pdf(968.31 KB)   Additional Information: full citation, abstract, references, citings, index terms

> This paper presents techniques for managing solid models in modern relational database management systems. Our goal is to enable support for traditional database operations (sorting, distance metrics, range queries, nearest neighbors, etc) on large databases of solid models. As part of this research, we have developed a number of novel storage and retrieval strategies that extend the state-of-the-art in database research as well as change the way in which solid modeling software developers an ...

Keywords: database clustering, database indexing, geometric reasoning, shape similarity, solid modeling

Results 1 - 3 of 3

	Type	Hits
1	BRS	519
2	IS&R	16129
3	BRS	58
4	BRS	319
5	BRS	521
6	IS&R	16181
7	BRS	58
8	BRS	21
9	BRS	51
10	BRS	30
11	IS&R	1454
12	BRS	5325
13	BRS	54
14	BRS	28
15	BRS	9
16	BRS	17
17	BRS	3044
18	BRS	5325
19	BRS	3681
20	BRS	374
21	BRS	11
22	BRS	811
23	BRS	3044
24	BRS	978
25	BRS	123
26	BRS	978
27	BRS	139
28	BRS	18
29	BRS	1
30	BRS	4
31	BRS	9

	Search Text
1	nearest ADJ neighbor SAME (class classif$8 catogor$8)
2	(707/$6).CCLS.
3	S1 and S2
4	nearest ADJ neighbor SAME (cluster$4)
5	nearest ADJ neighbor SAME (class classif$8 catogor$8)
6	(707/$6).CCLS.
7	S5 and S6
8	S4 and S7
9	S4 and S6
10	S9 not S7
11	(707/5).CCLS.
12	nearest ADJ neighbor
13	S11 and S12
14	S13 not (S7 S9)
15	nearest ADJ neighbor SAME (class classif$8 catogor$8 cluster$4) SAME (document text publlication) WITH (add$4 new delet$4 remov$4 chang$4)
16	nearest ADJ neighbor SAME (class classif$8 catogor$8 cluster$4) SAME (document text publlication) WITH (add$4 new delet$4 remov$4 chang$4)
17	707/3,5.ccls.
18	nearest ADJ neighbor
19	nearest ADJ neighbor
20	nearest ADJ neighbor.clm.
21	S17 and S20
22	(class classif$8 catogor$8 cluster$4) SAME (document text publlication) .clm.
23	S17
24	(class classif$8 catogor$8 cluster$4) SAME (document text publlication) .clm.
25	S17 and S24
26	(class classif$8 catogor$8 cluster$4) SAME (document text publlication).clm.
27	(class classif$8 catogor$8 cluster$4) SAME (document text publlication) WITH (add$4 new delet$4 remov$4 chang$4).clm.
28	S17 and S27
29	nearest ADJ neighbor SAME (class classif$8 catogor$8 cluster$4) SAME (document text publlication) WITH (add$4 new delet$4 remov$4 chang$4)
30	nearest ADJ neighbor SAME (class classif$8 catogor$8 cluster$4) SAME (document text publlication) SAME (add$4 new delet$4 remov$4

	chang$4)
31	nearest ADJ neighbor SAME (class classif$8 catogor$8 cluster$4) SAME (document text publlication)

	DBs	Time Stamp	Comments
1	USPAT	2006/01/10 10:21	
2	USPAT	2006/01/09 14:28	
3	USPAT	2006/01/09 14:28	
4	USPAT	2006/01/10 10:22	
5	USPAT	2006/01/10 15:36	
6	USPAT	2006/01/10 10:22	
7	USPAT	2006/01/10 10:22	
8	USPAT	2006/01/10 10:23	
9	USPAT	2006/01/10 10:23	
10	USPAT	2006/01/10 14:58	
11	USPAT	2006/01/10 14:58	
12	USPAT	2006/01/10 14:59	
13	USPAT	2006/01/10 14:59	
14	USPAT	2006/01/10 14:59	
15	USPAT	2006/01/10 15:46	
16	US-PGPUB	2006/01/10 15:52	
17	US-PGPUB	2006/01/10 15:36	
18	USPAT	2006/01/10 15:37	
19	US-PGPUB	2006/01/10 15:37	
20	US-PGPUB	2006/01/10 15:37	
21	US-PGPUB	2006/01/10 15:37	
22	USPAT	2006/01/10 15:49	
23	US-PGPUB	2006/01/10 15:49	
24	US-PGPUB	2006/01/10 15:50	
25	US-PGPUB	2006/01/10 15:51	
26	US-PGPUB	2006/01/10 15:51	
27	US-PGPUB	2006/01/10 15:51	
28	US-PGPUB	2006/01/10 15:51	
29	EPO; JPO; DERWENT; IBM_TDB	2006/01/10 15:53	
30	EPO; JPO; DERWENT; IBM_TDB	2006/01/12 13:42	
31	EPO; JPO; DERWENT; IBM_TDB	2006/01/12 13:47	

	Error Definition	Errors	Ref #
1			S1
2			S2
3			S3
4			S4
5			S5
6			S6
7			S7
8			S8
9			S9
10			S10
11			S11
12			S12
13			S13
14			S14
15			S15
16			S16
17			S17
18			S18
19			S19
20			S20
21			S21
22			S22
23			S23
24			S24
25			S25
26			S26
27			S27
28			S28
29			S29
30			S30
31			S31

	Type	Hits
32	BRS	4
33	BRS	4
34	BRS	5

	Search Text
32	nearest ADJ neighbor SAME (class classif$8 catogor$8 cluster$4) SAME (document text publlication) SAME (add$4 new delet$4 remov$4 chang$4)
33	S31 and S32
34	S31 not S32

	DBs	Time Stamp	Comments
32	EPO; JPO; DERWENT; IBM_TDB	2006/01/12 13:48	
33	EPO; JPO; DERWENT; IBM_TDB	2006/01/12 13:48	
34	EPO; JPO; DERWENT; IBM_TDB	2006/01/12 13:49	

	Error Definition	Errors	Ref #
32			S32
33			S33
34			S34

THE ACM DIGITAL LIBRARY

ⁱ᷃ Feedback Report a problem Satisfaction survey

Terms used
nearest neighbor cluster document add new delete remove insert change

Found **552** of **169,166**

Sort results by | relevance ▾

Display results | expanded form ▾

❧ Save results to a Binder
? Search Tips
☐ Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 20 of 200   Result page: **1** 2 3 4 5 6 7 8 9 10   next
Best 200 shown   Relevance scale ☐ ◩ ◪ ◼ ■

1 **Poly-logarithmic deterministic fully-dynamic algorithms for connectivity, minimum**
◈ **spanning tree, 2-edge, and biconnectivity**
   Jacob Holm, Kristian de Lichtenberg, Mikkel Thorup
   July 2001 **Journal of the ACM (JACM)**, Volume 48 Issue 4
   Publisher: ACM Press

   Full text available: 📄 pdf(378.20 KB)   Additional Information: full citation, abstract, references, citings, index terms

   Deterministic fully dynamic graph algorithms are presented for connectivity, minimum
   spanning tree, 2-edge connectivity, and biconnectivity. Assuming that we start with no
   edges in a graph with n vertices, the amortized operation costs are $O(\log^2 n)$ for
   connectivity, $O(\log^4 n)$ for minimum spanning forest, 2-edge connectivity, and $O(\log^5 n)$
   biconnectivity.

   Keywords: 2-edge connectivity, Biconnectivity, connectivity, dynamic graph algorithms,
   minimum spanning tree

2 Real-time shading
◈ Marc Olano, Kurt Akeley, John C. Hart, Wolfgang Heidrich, Michael McCool, Jason L. Mitchell,
   Randi Rost
   August 2004 **Proceedings of the conference on SIGGRAPH 2004 course notes GRAPH
              '04**
   Publisher: ACM Press
   Full text available: 📄 pdf(7.39 MB)   Additional Information: full citation, abstract

   Real-time procedural shading was once seen as a distant dream. When the first version of
   this course was offered four years ago, real-time shading was possible, but only with one-
   of-a-kind hardware or by combining the effects of tens to hundreds of rendering passes.
   Today, almost every new computer comes with graphics hardware capable of interactively
   executing shaders of thousands to tens of thousands of instructions. This course has been
   redesigned to address today's real-time shading capabili ...

3 Level set and PDE methods for computer graphics
◈ David Breen, Ron Fedkiw, Ken Museth, Stanley Osher, Guillermo Sapiro, Ross Whitaker
   August 2004 **Proceedings of the conference on SIGGRAPH 2004 course notes GRAPH
              '04**
   Publisher: ACM Press
   Full text available: 📄 pdf(17.07 MB)   Additional Information: full citation, abstract

   Level set methods, an important class of partial differential equation (PDE) methods,
   define dynamic surfaces implicitly as the level set (iso-surface) of a sampled, evolving nD
   function. The course begins with preparatory material that introduces the concept of using

partial differential equations to solve problems in computer graphics, geometric modeling and computer vision. This will include the structure and behavior of several different types of differential equations, e.g. the level set eq ...

4 Document expansion for speech retrieval

Amit Singhal, Fernando Pereira

August 1999 **Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(253.45 KB)   Additional Information: <u>full citation</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

5 Point-based computer graphics

Marc Alexa, Markus Gross, Mark Pauly, Hanspeter Pfister, Marc Stamminger, Matthias Zwicker

August 2004 **Proceedings of the conference on SIGGRAPH 2004 course notes GRAPH '04**

Publisher: ACM Press

Full text available: pdf(8.94 MB)   Additional Information: <u>full citation</u>, <u>abstract</u>

This course introduces points as a powerful and versatile graphics primitive. Speakers present their latest concepts for the acquisition, representation, modeling, processing, and rendering of point sampled geometry along with applications and research directions. We describe algorithms and discuss current problems and limitations, covering important aspects of point based graphics.

6 An optimal algorithm for approximate nearest neighbor searching fixed dimensions

Sunil Arya, David M. Mount, Nathan S. Netanyahu, Ruth Silverman, Angela Y. Wu

November 1998 **Journal of the ACM (JACM)**, Volume 45 Issue 6

Publisher: ACM Press

Full text available: pdf(287.94 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

Consider a set of S of n data points in real d-dimensional space, Rd, where distances are measured using any Minkowski metric. In nearest neighbor searching, we preprocess S into a data structure, so that given any query point $q \in$ Rd, is the closest point of S to q can be reported quickly. Given any po ...

Keywords: approximation algorithms, box-decomposition trees, closet-point queries, nearest neighbor searching, post-office problem, priority search

7 Searching in metric spaces by spatial approximation

Gonzalo Navarro

August 2002 **The VLDB Journal — The International Journal on Very Large Data Bases**, Volume 11 Issue 1

Publisher: Springer-Verlag New York, Inc.

Full text available: pdf(281.75 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>citings</u>, <u>index terms</u>

We propose a new data structure to search in metric spaces. A *metric space* is formed by a collection of objects and a *distance function* defined among them which satisfies the triangle inequality. The goal is, given a set of objects and a query, retrieve those objects close enough to the query. The complexity measure is the number of distances computed to achieve this goal. Our data structure, called *sa-tree* ("spatial approximation tree"), is based on approaching ...

Keywords: Multimedia databases, Similarity or proximity search, Spatial and multidimensional search, Spatial approximation tree

8 Interactive Editing Systems: Part II

Norman Meyrowitz, Andries van Dam
September 1982 **ACM Computing Surveys (CSUR)**, Volume 14 Issue 3
Publisher: ACM Press
Full text available: pdf(9.17 MB)   Additional Information: full citation, references, citings, index terms

9 External memory algorithms and data structures: dealing with **massive data**

Jeffrey Scott Vitter
June 2001 **ACM Computing Surveys (CSUR)**, Volume 33 Issue 2
Publisher: ACM Press
Full text available: pdf(828.46 KB)   Additional Information: full citation, abstract, references, citings, index terms

Data sets in large applications are often too massive to fit completely inside the computers internal memory. The resulting input/output communication (or I/O) between fast internal memory and slower external memory (such as disks) can be a major performance bottleneck. In this article we survey the state of the art in the design and analysis of external memory (or EM) algorithms and data structures, where the goal is to exploit locality in order to reduce the I/O costs. We consider a varie ...

Keywords: B-tree, I/O, batched, block, disk, dynamic, extendible hashing, external memory, hierarchical memory, multidimensional access methods, multilevel memory, online, out-of-core, secondary storage, sorting

10 Research sessions: clustering: Incremental and effective data summarization for dynamic hierarchical clustering

Samer Nassar, Jörg Sander, Corrine Cheng
June 2004 **Proceedings of the 2004 ACM SIGMOD international conference on Management of data**
Publisher: ACM Press
Full text available: pdf(235.15 KB)   Additional Information: full citation, abstract, references

Mining informative patterns from very large, dynamically changing databases poses numerous interesting challenges. Data summarizations (e.g., data bubbles) have been proposed to compress very large static databases into representative points suitable for subsequent effective hierarchical cluster analysis. In many real world applications, however, the databases dynamically change due to frequent insertions and deletions, possibly changing the data distribution and clustering structure over time. ...

Keywords: clustering, data summarization, incremental data bubbles

11 The elements of nature: interactive and realistic techniques

Oliver Deusen, David S. Ebert, Ron Fedkiw, F. Kenton Musgrave, Przemyslaw Prusinkiewicz, Doug Roble, Jos Stam, Jerry Tessendorf
August 2004 **Proceedings of the conference on SIGGRAPH 2004 course notes GRAPH '04**
Publisher: ACM Press
Full text available: pdf(17.65 MB)   Additional Information: full citation, abstract

This updated course on simulating natural phenomena will cover the latest research and production techniques for simulating most of the elements of nature. The presenters will provide movie production, interactive simulation, and research perspectives on the difficult task of photorealistic modeling, rendering, and animation of natural phenomena. The course offers a nice balance of the latest interactive graphics hardware-based simulation techniques and the latest physics-based simulation techni ...

12 Index-driven similarity search in metric spaces

Gisli R. Hjaltason, Hanan Samet
December 2003 **ACM Transactions on Database Systems (TODS)**, Volume 28 Issue 4

Full text available: pdf(650.64 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

Similarity search is a very important operation in multimedia databases and other database applications involving complex objects, and involves finding objects in a data set S similar to a query object q, based on some similarity measure. In this article, we focus on methods for similarity search that make the general assumption that similarity is represented with a distance metric d. Existing methods for handling similarity search in this setting typically fall into one of ...

Keywords: Hiearchical metric data structures, distance-based indexing, nearest neighbor queries, range queries, ranking, similarity searching

13 The string edit distance matching problem with moves

Graham Cormode, S. Muthukrishnan
January 2002 **Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms**
Full text available: pdf(1.13 MB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>

The edit distance between two strings S and R is defined to be the minimum number of character inserts, deletes and changes needed to convert R to S. Given a text string t of length n, and a pattern string p of length m, informally, the string edit distance matching problem is to compute the smallest edit distance between p and substrings of t. A well known dynamic programming algorithm takes time $O(nm)$ to solve ...

14 Link-based similarity: LSH forest: self-tuning indexes for similarity search

Mayank Bawa, Tyson Condie, Prasanna Ganesan
May 2005 **Proceedings of the 14th international conference on World Wide Web**
Full text available: pdf(247.91 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>index terms</u>

We consider the problem of indexing high-dimensional data for answering (approximate) similarity-search queries. Similarity indexes prove to be important in a wide variety of settings: Web search engines desire fast, parallel, main-memory-based indexes for similarity search on text data; database systems desire disk-based similarity indexes for high-dimensional data, including text and images; peer-to-peer systems desire distributed similarity indexes with low communication cost. We propose an i ...

Keywords: peer-to-peer (P2P), similarity indexes

15 Distance browsing in spatial databases

Gísli R. Hjaltason, Hanan Samet
June 1999 **ACM Transactions on Database Systems (TODS)**, Volume 24 Issue 2

Full text available: pdf(460.81 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

We compare two different techniques for browsing through a collection of spatial objects stored in an R-tree spatial data structure on the basis of their distances from an arbitrary spatial query object. The conventional approach is one that makes use of a k-nearest neighbor algorithm where k is known prior to the invocation of the algorithm. Thus if m < k neighbors are needed, the k-nearest neighbor alg ...

Keywords: R-trees, distance browsing, hiearchical spatial data structures, nearest neighbors, ranking

16

Crowd and group animation

Daniel Thalmann, Christophe Hery, Seth Lippman, Hiromi Ono, Stephen Regelous, Douglas Sutton
August 2004 **Proceedings of the conference on SIGGRAPH 2004 course notes GRAPH '04**
Publisher: ACM Press
Full text available: pdf(20.19 MB)   Additional Information: <u>full citation</u>, <u>abstract</u>

A continuous challenge for special effects in movies is the production of realistic virtual crowds, in terms of rendering and behavior. This course will present state-of-the-art techniques and methods. The course will explain in details the different approaches to create virtual crowds: particle systems with flocking techniques using attraction and repulsion forces, copy and pasting techniques, agent-based methods. The architecture of software tools will be presented including the MASSIVE softwa ...

17 RCV1: A New Benchmark Collection for Text Categorization Research

David D. Lewis, Yiming Yang, Tony G. Rose, Fan Li
December 2004 **The Journal of Machine Learning Research**, Volume 5
Publisher: MIT Press

Full text available: pdf(628.29 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>citings</u>, <u>index terms</u>

Reuters Corpus Volume I (RCV1) is an archive of over 800,000 manually categorized newswire stories recently made available by Reuters, Ltd. for research purposes. Use of this data for research on text categorization requires a detailed understanding of the real world constraints under which the data was produced. Drawing on interviews with Reuters personnel and access to Reuters documentation, we describe the coding policy and quality control procedures used in producing the RCV1 data, the inten ...

18 Level II technical support in a distributed computing environment

Tim Leehane
September 1996 **Proceedings of the 24th annual ACM SIGUCCS conference on User services**
Publisher: ACM Press

Full text available: pdf(5.73 MB)   Additional Information: <u>full citation</u>, <u>references</u>, <u>index terms</u>

19 Automated hoarding for mobile computers

Geoffrey H. Kuenning, Gerald J. Popek
October 1997 **ACM SIGOPS Operating Systems Review , Proceedings of the sixteenth ACM symposium on Operating systems principles SOSP '97**, Volume 31 Issue 5
Publisher: ACM Press

Full text available: pdf(2.05 MB)   Additional Information: <u>full citation</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

20 Similarity queries I: Robust and efficient fuzzy match for online data cleaning

Surajit Chaudhuri, Kris Ganjam, Venkatesh Ganti, Rajeev Motwani
June 2003 **Proceedings of the 2003 ACM SIGMOD international conference on Management of data**
Publisher: ACM Press

Full text available: pdf(271.47 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

To ensure high data quality, data warehouses must validate and cleanse incoming data tuples from external sources. In many situations, clean tuples must match acceptable tuples in *reference tables*. For example, product name and description fields in a sales record from a distributor must match the pre-recorded name and description fields in a product reference relation.A significant challenge in such a scenario is to implement an efficient and accurate fuzzy match operation that can effec ...

ⁱ⁶ Feedback Report a problem Satisfaction survey

An optimal algorithm for approximate nearest neighbor searching fixed dimensions

Full text   📄Pdf (288 KB)

Authors   Sunil Arya   Hong Kong Univ. of Science of Technology, Kowloon, Hong Kong
         David M. Mount   Univ. of Maryland, College Park
         Nathan S. Netanyahu Univ. of Maryland, College Park; and NASA, Greenbelt, MD
         Ruth Silverman   Univ. of the District of Columbia, Washington, D.C.; and Univ. of Maryland, College Park
         Angela Y. Wu   American Univ., Washington, D.C.

Additional Information: abstract references citings index terms collaborative colleagues peer to peer

Tools and Actions: Discussions   Find similar Articles   Review this Article
           Save this Article to a Binder   Display Formats: BibTex EndNote ACM Ref

DOI Bookmark: Use this link to bookmark this Article: http://doi.acm.org/10.1145/293347.293348
          What is a DOI?

↑ ABSTRACT

Consider a set of S of n data points in real d-dimensional space, Rd, where distances are measured using any Minkowski metric. In nearest neighbor searching, we preprocess S into a data structure, so that given any query point q $\in$ Rd, is the closest point of S to q can be reported quickly. Given any positive real ϵ, data point p is a $(1 + \epsilon)$-approximate nearest neighbor of q if its distance from q is within a factor of $(1 + \epsilon)$ of the distance to the true nearest neighbor. We show that it is possible to preprocess a set of n points in Rd in $O(dn \log n)$ time and $O(dn)$ space, so that given a query point q $\in$ Rd, and $\epsilon > 0$, a $(1 + \epsilon)$-approximate nearest neighbor of q can be computed in $O(c_{d, \epsilon} \log n)$ time, where $c_{d,\epsilon} \leq d \, 1 + 6d/\epsilon;d$ is a factor depending only on dimension and ϵ. In general, we show that given an integer $k \geq 1$, $(1 + \epsilon)$-approximations to the k nearest neighbors of q can be computed in additional $O(kd \log n)$ time.

↑ REFERENCES

Note: OCR errors may be found in this Reference List extracted from the full text article. ACM has opted to expose the complete List rather than only correct and linked references.

1 Pankaj K. Agarwal , Jiří Matoušek, Ray shooting and parametric search, SIAM Journal on Computing, v.22 n.4, p.794-806, Aug. 1993

2 ARYA, S., AND MOUNT, D.M. 1993a. Algorithms for fast vector quantization. In Proceedings of the DCC'93: Data Compression Conference. J. A. Storer and M. Cohn, eds. IEEE Press, New York, pp. 381-390.

3 Sunil Arya , David M. Mount, Approximate nearest neighbor queries in fixed dimensions, Proceedings of the fourth annual ACM-SIAM Symposium on Discrete algorithms, p.271-280, January

25-27, 1993, Austin, Texas, United States

4   Sunil Arya , David M. Mount, Approximate range searching, Proceedings of the eleventh annual symposium on Computational geometry, p.172-181, June 05-07, 1995, Vancouver, British Columbia, Canada

5   Sunil Arya , David M. Mount , Onuttom Narayan, Accounting for boundary effects in nearest neighbor searching, Proceedings of the eleventh annual symposium on Computational geometry, p.336-344, June 05-07, 1995, Vancouver, British Columbia, Canada

6   Sunil Arya , David M. Mount , Nathan S. Netanyahu , Ruth Silverman , Angela Wu, An optimal algorithm for approximate nearest neighbor searching, Proceedings of the fifth annual ACM-SIAM symposium on Discrete algorithms, p.573-582, January 23-25, 1994, Arlington, Virginia, United States

7   BEI, C.-D., AND GRAY, R. M. 1985. An improvement of the minimum distortion encoding algorithm for vector quantization. IEEE Trans. Commun. 33, 1132-1133.

8   Jon Louis Bentley , Bruce W. Weide , Andrew C. Yao, Optimal Expected-Time Algorithms for Closest Point Problems, ACM Transactions on Mathematical Software (TOMS), v.6 n.4, p.563-580, Dec. 1980

9   Stefan Berchtold , Christian Böhm , Daniel A. Keim , Hans-Peter Kriegel, A cost model for nearest neighbor search in high-dimensional data space, Proceedings of the sixteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems, p.78-86, May 11-15, 1997, Tucson, Arizona, United States

10   Stefan Berchtold , Daniel A. Keim , Hans-Peter Kriegel, The X-tree: An Index Structure for High-Dimensional Data, Proceedings of the 22th International Conference on Very Large Data Bases, p.28-39, September 03-06, 1996

11   Marshall Bern, Approximate closest-point queries in high dimensions, Information Processing Letters, v.45 n.2, p.95-99, Feb. 26, 1993

12   Marshall W. Bern , David Eppstein , Shang-Hua Teng, Parallel Construction of Quadtrees and Quality Triangulations, Proceedings of the Third Workshop on Algorithms and Data Structures, p.188-199, August 11-13, 1993

13   Sergei N. Bespamyatnikh, An optimal algorithm for closest pair maintenance (extended abstract), Proceedings of the eleventh annual symposium on Computational geometry, p.152-161, June 05-07, 1995, Vancouver, British Columbia, Canada

14   Paul B. Callahan , S. Rao Kosaraju, A decomposition of multi-dimensional point-sets with applications to *k*-nearest-neighbors and *n*-body potential fields (preliminary version), Proceedings of the twenty-fourth annual ACM symposium on Theory of computing, p.546-556, May 04-06, 1992, Victoria, British Columbia, Canada

15   Paul B. Callahan , S. Rao Kosaraju, Algorithms for dynamic closest pair and n-body potential fields, Proceedings of the sixth annual ACM-SIAM symposium on Discrete algorithms, p.263-272, January 22-24, 1995, San Francisco, California, United States

16   Timothy M. Chan, Approximate nearest neighbor queries revisited, Proceedings of the thirteenth annual symposium on Computational geometry, p.352-358, June 04-06, 1997, Nice, France

17   CHAZELLE, B. 1982. A theorem on polygon cutting with applications. In Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science. IEEE Computer Society Press, Los Alamitos, Calif., pp. 339-349.

18   CLARKSON, K.L. 1983. Fast algorithms for the all nearest neighbors problem. In Proceedings of the 24th Annual IEEE Symposium on the Foundations of Computer Science. IEEE Computer Society

Press, Los Alamitos, Calif., pp. 226-232.

19 Kenneth L. Clarkson, A randomized algorithm for closest-point queries, SIAM Journal on Computing, v.17 n.4, p.830-847, August 1988

20 Kenneth L. Clarkson, An algorithm for approximate closest-point queries, Proceedings of the tenth annual symposium on Computational geometry, p.160-164, June 06-08, 1994, Stony Brook, New York, United States

21 John Gerald Cleary, Analysis of an Algorithm for Finding Nearest Neighbors in Euclidean Space, ACM Transactions on Mathematical Software (TOMS), v.5 n.2, p.183-192, June 1979

22 Thomas T. Cormen , Charles E. Leiserson , Ronald L. Rivest, Introduction to algorithms, MIT Press, Cambridge, MA, 1990

23 Scott Cost , Steven Salzberg, A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features, Machine Learning, v.10 n.1, p.57-78, Jan. 1993

24 COVER, T. M., AND HART, P.E. 1967. Nearest neighbor pattern classification. IEEE Trans. Info. Theory 13, 57-67.

25 Mark de Berg , Marc van Kreveld , Mark Overmars , Otfried Schwarzkopf, Computational geometry: algorithms and applications, Springer-Verlag New York, Inc., Secaucus, NJ, 1997

26 DEERWESTER, S., DUMALS, S. T., FURNAS, G. W., LANDAUER, T. K., AND HARSHMAN, R. 1990. Indexing by latend semantic analysis. J. Amer. Soc. Inf. Sci. 41,391-407.

27 DEVaOYE, L., AND WAGNEa, T.J. 1982. Nearest neighbor methods in discrimination. In Handbook of Statistics, vol. 2. P. R. Krishnaiah and L. N. Kanal, eds. North-Holland, Amsterdam, The Netherlands.

28 DUDA, R. O., AND HAST, P.E. 1978. Pattern Classification and Scene Analysis. Wiley, New York.

29 Herbert Edelsbrunner, Algorithms in combinatorial geometry, Springer-Verlag New York, Inc., New York, NY, 1987

30 FARVARDIN, N., AND MODESTINO, J.W. 1985. Rate-distortion performance of DPCM schemes for sutoregressive sources. IEEE Trans. Inf. Theory 31, 3 (May), 402-418.

31 Usama M. Fayyad , Gregory Piatetsky-Shapiro , Padhraic Smyth , Ramasamy Uthurusamy, Advances in knowledge discovery and data mining, American Association for Artificial Intelligence, Menlo Park, CA, 1996

32 Tomás Feder , Daniel Greene, Optimal algorithms for approximate clustering, Proceedings of the twentieth annual ACM symposium on Theory of computing, p.434-444, May 02-04, 1988, Chicago, Illinois, United States

33 Myron Flickner , Harpreet Sawhney , Wayne Niblack , Jonathan Ashley , Qian Huang , Byron Dom , Monika Gorkani , Jim Hafner , Denis Lee , Dragutin Petkovic , David Steele , Peter Yanker, Query by Image and Video Content: The QBIC System, Computer, v.28 n.9, p.23-32, September 1995

34 FREDERICKSON, G.N. 1985. Data structures for on-line updating of minimum spanning trees, with applications. SIAM J. Comput. 14, 781-798.

35 Greg N. Frederickson, A data structure for dynamically maintaining rooted trees, Proceedings of the fourth annual ACM-SIAM Symposium on Discrete algorithms, p.175-184, January 25-27, 1993, Austin, Texas, United States

36 Michael L. Fredman , Robert Endre Tarjan, Fibonacci heaps and their uses in improved network

optimization algorithms, Journal of the ACM (JACM), v.34 n.3, p.596-615, July 1987

37   Jerome H. Freidman , Jon Louis Bentley , Raphael Ari Finkel, An Algorithm for Finding Best Matches in Logarithmic Expected Time, ACM Transactions on Mathematical Software (TOMS), v.3 n.3, p.209-226, Sept. 1977

38   FRIEDMAN, J. H., BASKETT, F., AND SHUSTEK, L. J. 1975. An algorithm for finding nearest neighbors. IEEE Trans. Comput. C-24, 10, 1000-1006.

39   Igal Galperin , Ronald L. Rivest, Scapegoat trees, Proceedings of the fourth annual ACM-SIAM Symposium on Discrete algorithms, p.165-174, January 25-27, 1993, Austin, Texas, United States

40   Allen Gersho , Robert M. Gray, Vector quantization and signal compression, Kluwer Academic Publishers, Norwell, MA, 1991

41   L. Guan , M. Kamel, Equal-average hyperplane partitioning method for vector quantization of image data, Pattern Recognition Letters, v.13 n.10, p.693-699, Oct. 1992

42   Piotr Indyk , Rajeev Motwani, Approximate nearest neighbors: towards removing the curse of dimensionality, Proceedings of the thirtieth annual ACM symposium on Theory of computing, p.604-613, May 24-26, 1998, Dallas, Texas, United States

43   Jon M. Kleinberg, Two algorithms for nearest-neighbor search in high dimensions, Proceedings of the twenty-ninth annual ACM symposium on Theory of computing, p.599-608, May 04-06, 1997, El Paso, Texas, United States

44   Eyal Kushilevitz , Rafail Ostrovsky , Yuval Rabani, Efficient search for approximate nearest neighbor in high dimensional spaces, Proceedings of the thirtieth annual ACM symposium on Theory of computing, p.614-623, May 24-26, 1998, Dallas, Texas, United States

45   LEE, C.-H., AND CHEN, L.-H. 1994. Fast closest codeword search algorithm for vector quantisation. IEEE Proc. Vis. Image Sig. Proc. 141, 143-148.

46   King Ip Lin , H. V. Jagadish , Christos Faloutsos, The TV-tree: an index structure for high-dimensional data, The VLDB Journal — The International Journal on Very Large Data Bases, v.3 n.4, October 1994

47   S. Meiser, Point location in arrangements of hyperplanes, Information and Computation, v.106 n.2, p.286-303, Oct. 1993

48   MOUNT, D. M., NETANYAHU, N., SILVERMAN, R., AND WU, A. Y. 1995. Chromatic nearest neighbor searching: A query sensitive approach. In Proceedings of the 7th Canadian Conference on Computer Geometry. (Quebec City, Que., Canada, Aug. 10-13). pp. 261-266.

49   Franco P. Preparata , Michael I. Shamos, Computational geometry: an introduction, Springer-Verlag New York, Inc., New York, NY, 1985

50   RIVEST, R.L. 1974. On the optimality of Elais's algorithm for performing best-match searches. In Information Processing. North-Holland Publishing Company, Amsterdam, The Netherlands, pp. 678-681.

51   Nick Roussopoulos , Stephen Kelley , Frédéric Vincent, Nearest neighbor queries, Proceedings of the 1995 ACM SIGMOD international conference on Management of data, p.71-79, May 22-25, 1995, San Jose, California, United States

52   Hanan Samet, The design and analysis of spatial data structures, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, 1990

53   SCHWARZ, C., SMID, M., AND SNOEYINK, J. 1994. An optimal algorithm for the on-line closest-pair problem. Algorithmica 12, 18-29.

54   Daniel D. Sleator , Robert Endre Tarjan, A data structure for dynamic trees, Journal of Computer and System Sciences, v.26 n.3, p.362-391, June 1983

55   SPROULL, R.L. 1991. Refinements to nearest-neighbor searching. Algorithmica 6, 579-589.

56   P. M. Vaidya, An O(n log n) algorithm for the all-nearest-neighbors problem, Discrete & Computational Geometry, v.4 n.2, p.101-115, Jan. 1989

57   David A. White , Ramesh Jain, Similarity Indexing with the SS-tree, Proceedings of the Twelfth International Conference on Data Engineering, p.516-523, February 26-March 01, 1996

58   A C Yao , F F Yao, A general approach to d-dimensional geometric queries, Proceedings of the seventeenth annual ACM symposium on Theory of computing, p.163-168, May 06-08, 1985, Providence, Rhode Island, United States

↑ CITINGS *46*

Robert Krauthgamer , James R. Lee, Navigating nets: simple algorithms for proximity search, Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms, January 11-14, 2004, New Orleans, Louisiana

Sariel Har-Peled , Manor Mendel, Fast construction of nets in low dimensional metrics, and their applications, Proceedings of the twenty-first annual symposium on Computational geometry, June 06-08, 2005, Pisa, Italy

Sylvain Lefebvre , Hugues Hoppe, Parallel controllable texture synthesis, ACM Transactions on Graphics (TOG), v.24 n.3, July 2005

Martin Peternell , Tibor Steiner, A geometric idea to solve the eikonal equation, Proceedings of the 21st spring conference on Computer graphics, May 12-14, 2005, Budmerice, Slovakia

Xin Tong , Jingdan Zhang , Ligang Liu , Xi Wang , Baining Guo , Heung-Yeung Shum, Synthesis of bidirectional texture functions on arbitrary surfaces, ACM Transactions on Graphics (TOG), v.21 n.3, July 2002

Lili Ju , Qiang Du , Max Gunzburger, Probabilistic methods for centroidal Voronoi tessellations and their parallel implementations, Parallel Computing, v.28 n.10, p.1477-1500, October 2002

A. Corral , Y. Manolopoulos , Y. Theodoridis , M. Vassilakopoulos, Algorithms for processing K-closest-pair queries in spatial databases, Data & Knowledge Engineering, v.49 n.1, p.67-104, 1 April 2004

Joachim Gudmundsson , Marc van Kreveld , Bettina Speckmann, Efficient detection of motion patterns in spatio-temporal data sets, Proceedings of the 12th annual ACM international workshop on Geographic information systems, November 12-13, 2004, Washington DC, USA

Philip David , Daniel Dementhon , Ramani Duraiswami , Hanan Samet, SoftPOSIT: Simultaneous Pose and Correspondence Determination, International Journal of Computer Vision, v.59 n.3, p.259-284, September-October 2004

Stephan Volmer, Buoy indexing of metric feature spaces for fast approximate image queries, Proceedings of the sixth Eurographics workshop on Multimedia 2001, p.131-140, September 08-09, 2001, Manchester, UK

Aaron Hertzmann , Charles E. Jacobs , Nuria Oliver , Brian Curless , David H. Salesin, Image analogies, Proceedings of the 28th annual conference on Computer graphics and interactive techniques, p.327-340, August 2001

Lin Liang , Ce Liu , Ying-Qing Xu , Baining Guo , Heung-Yeung Shum, Real-time texture synthesis by patch-based sampling, ACM Transactions on Graphics (TOG), v.20 n.3, p.127-150, July 2001

Timothy M. Chan, Closest-point problems simplified on the RAM, Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms, p.472-473, January 06-08, 2002, San Francisco, California

Sunil Arya , Ho-Yam Addy Fu, Expected-case complexity of approximate nearest neighbor searching, Proceedings of the eleventh annual ACM-SIAM symposium on Discrete algorithms, p.379-388, January 09-11, 2000, San Francisco, California, United States

Gisli R. Hjaltason , Hanan Samet, Index-driven similarity search in metric spaces, ACM Transactions on Database Systems (TODS), v.28 n.4, p.517-580, December 2003

Like Gao , Min Wang , X. Sean Wang, Quality-driven evaluation of trigger conditions on streaming time series, Proceedings of the 2005 ACM symposium on Applied computing, March 13-17, 2005, Santa Fe, New Mexico

David G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, v.60 n.2, p.91-110, November 2004

Daniel Sýkora , Jan Buriánek , Jiří Žára, Unsupervised colorization of black-and-white cartoons, Proceedings of the 3rd international symposium on Non-photorealistic animation and rendering, June 07-09, 2004, Annecy, France

Guang-Ho Cha, Bitmap indexing method for complex similarity queries with relevance feedback, Proceedings of the first ACM international workshop on Multimedia databases, November 07-07, 2003, New Orleans, LA, USA

Sunil Arya , Theocharis Malamatos, Linear-size approximate voronoi diagrams, Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms, p.147-155, January 06-08, 2002, San Francisco, California

Sunil Arya , Theocharis Malamatos , David M. Mount, Space-efficient approximate Voronoi diagrams, Proceedings of the thiry-fourth annual ACM symposium on Theory of computing, May 19-21, 2002, Montreal, Quebec, Canada

George ElKoura , Karan Singh, Handrix: animating the human hand, Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, July 26-27, 2003, San Diego, California

Niloy J. Mitra , Natasha Gelfand , Helmut Pottmann , Leonidas Guibas, Registration of point cloud data from a geometric optimization perspective, Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing, July 08-10, 2004, Nice, France

Ting Liu , Ke Yang , Andrew W. Moore, The IOC algorithm: efficient many-class non-parametric classification for high-dimensional data, Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining, August 22-25, 2004, Seattle, WA, USA

H. V. Jagadish , Beng Chin Ooi , Kian-Lee Tan , Cui Yu , Rui Zhang, iDistance: An adaptive B^{+}-tree based indexing method for nearest neighbor search, ACM Transactions on Database Systems (TODS), v.30 n.2, p.364-397, June 2005

Paolo Liberatore, Monotonic reductions, representative equivalence, and compilation of intractable problems, Journal of the ACM (JACM), v.48 n.6, p.1091-1125, November 2001

Dimitris Papadias , Yufei Tao , Kyriakos Mouratidis , Chun Kit Hui, Aggregate nearest neighbor queries in spatial databases, ACM Transactions on Database Systems (TODS), v.30 n.2, p.529-576, June 2005

Katsu Yamane , James J. Kuffner , Jessica K. Hodgins, Synthesizing animations of human manipulation tasks, ACM Transactions on Graphics (TOG), v.23 n.3, August 2004

Timothy M. Chan, Polynomial-time approximation schemes for packing and piercing fat objects, Journal of Algorithms, v.46 n.2, p.178-189, February 2003

Siu-Wing Cheng , Tamal K. Dey , Edgar A. Ramos, Manifold reconstruction from point samples, Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms, January 23-25, 2005, Vancouver, British Columbia

Gísli R. Hjaltason , Hanan Samet, Distance browsing in spatial databases, ACM Transactions on Database Systems (TODS), v.24 n.2, p.265-318, June 1999

Tapas Kanungo , David M. Mount , Nathan S. Netanyahu , Christine Piatko , Ruth Silverman , Angela Y. Wu, The analysis of a simple k-means clustering algorithm, Proceedings of the sixteenth annual symposium on Computational geometry, p.100-109, June 12-14, 2000, Clear Water Bay, Kowloon, Hong Kong

Stefan Funke , Edgar A. Ramos, Smooth-surface reconstruction in near-linear time, Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms, p.781-790, January 06-08, 2002, San Francisco, California

Niloy J. Mitra , An Nguyen, Estimating surface normals in noisy point cloud data, Proceedings of the nineteenth conference on Computational geometry, June 08-10, 2003, San Diego, California, USA

Tapas Kanungo , David M. Mount , Nathan S. Netanyahu , Christine D. Piatko , Ruth Silverman , Angela Y. Wu, An Efficient k-Means Clustering Algorithm: Analysis and Implementation, IEEE Transactions on Pattern Analysis and Machine Intelligence, v.24 n.7, p.881-892, July 2002

Sunil Arya , Theocharis Malamatos , David M. Mount, Space-time tradeoffs for approximate spherical range counting, Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms, January 23-25, 2005, Vancouver, British Columbia

Jie Gao , Leonidas J. Guibas , An Nguyen, Deformable spanners and applications, Proceedings of the twentieth annual symposium on Computational geometry, June 08-11, 2004, Brooklyn, New York, USA

Giuseppe Amato , Fausto Rabitti , Pasquale Savino , Pavel Zezula, Region proximity in metric spaces and its use for approximate similarity search, ACM Transactions on Information Systems (TOIS), v.21 n.2, p.192-227, April 2003

Sariel Har-Peled , Soham Mazumdar, On coresets for k-means and k-median clustering, Proceedings of the thirty-sixth annual ACM symposium on Theory of computing, June 13-16, 2004, Chicago, IL, USA

Aristides Gionis , Alexander Hinneburg , Spiros Papadimitriou , Panayiotis Tsaparas, Dimension induced clustering, Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, August 21-24, 2005, Chicago, Illinois, USA

David Eppstein , Michael T. Goodrich , Jonathan Z. Sun, The skip quadtree: a simple dynamic data structure for multidimensional data, Proceedings of the twenty-first annual symposium on Computational geometry, June 06-08, 2005, Pisa, Italy

Dong-Ho Lee , Hyoung-Joo Kim, An Efficient Technique for Nearest-Neighbor Query Processing on the SPY-TEC, IEEE Transactions on Knowledge and Data Engineering, v.15 n.6, p.1472-1486, November 2003

Gísli R. Hjaltason , Hanan Samet, Properties of Embedding Methods for Similarity Searching in Metric Spaces, IEEE Transactions on Pattern Analysis and Machine Intelligence, v.25 n.5, p.530-549, May 2003

A. M. Roumani , D. B. Skillicorn, Mobile services discovery and selection in the publish/subscribe paradigm, Proceedings of the 2004 conference of the Centre for Advanced Studies conference on

Collaborative research, p.163-173, October 04-07, 2004, Markham, Ontario, Canada

Mircea Nicolescu , Gérard Medioni, Layered 4D Representation and Voting for Grouping from Motion, IEEE Transactions on Pattern Analysis and Machine Intelligence, v.25 n.4, p.492-501, April 2003

Pabitra Mitra , C. A. Murthy , Sankar K. Pal, Density-Based Multiscale Data Condensation, IEEE Transactions on Pattern Analysis and Machine Intelligence, v.24 n.6, p.734-747, June 2002

↑ INDEX TERMS

Primary Classification:
  E. Data

Additional Classification:
  F. Theory of Computation
  ↳ **F.2** ANALYSIS OF ALGORITHMS AND PROBLEM COMPLEXITY

  H. Information Systems
  ↳ **H.3** INFORMATION STORAGE AND RETRIEVAL

General Terms:
Algorithms, Theory

Keywords:
approximation algorithms, box-decomposition trees, closet-point queries, nearest neighbor searching, post-office problem, priority search

↑ Collaborative Colleagues:

Sunil Arya:	Siu-Wing Cheng	Kurt Mehlhorn	Angela Y. Wu	
	Gautam Das	David M. Mount		
	Larry S. Davis	Onuttom Narayan		
	Daniel DeMenthon	Nathan S. Netanyahu		
	Ho-Yam Addy Fu	H. Ramesh		
	Mordecai J. Golin	Jeffrey S. Salowe		
	Ramesh Hariharan	Ruth Silverman		
	V. S. Anil Kumar	Michiel Smid		
	Theocharis	Michiel H. M. Smid		
	Malamatos	Angela Wu		
	Peter Meer			
David M. Mount:	Esther M. Arkin	D. Yu. Grigoryev	Joseph S. B.	Alan Saalfeld
	Sunil Arya	Ramesh Hariharan	Mitchell	Jeffrey S.
	Fatma Betul	Tapas Kanungo	Michael Murphy	Salowe
	Atalay-Satoglu	Thomas Kao	Giri Narasimhan	Ruth
	László Babai	Thomas C. Kao	Onuttom Narayan	Silverman
	Sharat Chandran	Samir Khuller	Nathan S.	Steven
	Moses Charikar	T. Y. Kong	Netanyahu	Skiena
	Siu-Wing Cheng	Jacqueline Le Moigne	Christos H.	Michiel Smid
	Gautam Das	Theocharis Malamatos	Papadimitriou	Clifford Stein
	Carl W. Gable	Songrit	Christine Piatko	Subhash Suri
	Subir Kumar Ghosh	Maneewongvatana	Christine D. Piatko	Angela Wu
	Michael T. Goodrich	Jiří Matoušek	Fan-Tao Pu	Angela Y. Wu
			Kathleen Romanik	
			A. W. Roscoe	
Nathan S. Netanyahu:	Amihood Amir	Walter B. Ligon III	Keith Wichmann	
	Sunil Arya	Jiří Matoušek	Angela Wu	
	William J. Campbell	David M. Mount	Angela Y. Wu	

	Robert F. Cromp	Christine Piatko
	Gyorgy Fekete	Christine D. Piatko
	Tapas Kanungo	Kathleen Romanik
	Reuven Kashi	Nicholas M. Short
	Daniel Keim	Ruth Silverman
	Jacqueline Le Moigne	James C. Tilton
	Jacqueline J. LeMoigne	Markus Wawryniuk
Ruth Silverman:	Sunil Arya	Angela Y. Wu
	Tapas Kanungo	
	H. Levinson	
	David Mount	
	David M. Mount	
	Nathan S. Netanyahu	
	Christine Piatko	
	Christine D. Piatko	
	Kathleen Romanik	
	Angela Wu	
Angela Y. Wu:	Sunil Arya	Azriel Rosenfeld
	S. K. Bhaskar	Ruth Silverman
	Tapas Kanungo	Michael Werman
	T. Yung Kong	
	Robert A. Melter	
	David M. Mount	
	Nathan S. Netanyahu	
	Christine Piatko	
	Christine D. Piatko	
	Kathleen Romanik	

↑ Peer to Peer - Readers of this Article have also read:

- Data structures for quadtree approximation and compression
  Communications of the ACM 28, 9
  Hanan Samet

- A hierarchical single-key-lock access control using the Chinese remainder theorem
  Proceedings of the 1992 ACM/SIGAPP Symposium on Applied computing
  Kim S. Lee , Huizhu Lu , D. D. Fisher

- Putting innovation to work: adoption strategies for multimedia communication systems
  Communications of the ACM 34, 12
  Ellen Francik , Susan Ehrlich Rudman , Donna Cooper , Stephen Levine

- The GemStone object database management system
  Communications of the ACM 34, 10
  Paul Butterworth , Allen Otis , Jacob Stein

- An intelligent component database for behavioral synthesis
  Proceedings of the 27th ACM/IEEE conference on Design automation
  Gwo-Dong Chen , Daniel D. Gajski

P&RTAL

USPTO

Search: ◉ The ACM Digital Library   ○ The Guide

+nearest +neighbor +cluster +document +(precision recall)

SEARCH

ᵣ Feedback Report a problem Satisfaction survey

Terms used
nearest neighbor cluster document precision recall

Found **312** of **169,166**

Sort results by   [relevance ▼]

Display results   [expanded form ▼]

🔖 Save results to a Binder

⁉ Search Tips

☐ Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 20 of 200   Result page: **1** 2 3 4 5 6 7 8 9 10   next
Best 200 shown   Relevance scale ☐▢▣▤▦

1   **IR-7 (information retrieval): natural language processing for IR: Stemming and lemmatization in the clustering of finnish text documents**   ▦

Tuomo Korenius, Jorma Laurikkala, Kalervo Järvelin, Martti Juhola
November 2004 **Proceedings of the thirteenth ACM international conference on Information and knowledge management CIKM '04**
Publisher: ACM Press
Full text available: 🗎 pdf(238.51 KB)   Additional Information: full citation, abstract, references, index terms

Stemming and lemmatization were compared in the clustering of Finnish text documents. Since Finnish is a highly inflectional and agglutinative language, we hypothesized that lemmatization, involving splitting of the compound words, would be more appropriate normalization approach than the straightforward stemming. The relevance of the documents were evaluated with a four-point relevance assessment scale, which was collapsed into binary one by considering all the relevant and only the highly r ...

Keywords: clustering, lemmatization, normalization, stemming

2   **Shape-based retrieval and analysis of 3D models**   ▦

Thomas Funkhouser, Michael Kazhdan
August 2004 **Proceedings of the conference on SIGGRAPH 2004 course notes GRAPH '04**
Publisher: ACM Press
Full text available: 🗎 pdf(12.56 MB)   Additional Information: full citation, abstract

Large repositories of 3D data are rapidly becoming available in several fields, including mechanical CAD, molecular biology, and computer graphics. As the number of 3D models grows, there is an increasing need for computer algorithms to help people find the interesting ones and discover relationships between them. Unfortunately, traditional text-based search techniques are not always effective for 3D models, especially when queries are geometric in nature (e.g., find me objects that fit into thi ...

3   **The cluster hypothesis revisited**   ▦

Ellen M. Voorhees
June 1985 **Proceedings of the 8th annual international ACM SIGIR conference on Research and development in information retrieval**
Publisher: ACM Press
Full text available: 🗎 pdf(559.94 KB)   Additional Information: full citation, abstract, references, citings

A new means of evaluating the cluster hypothesis is introduced and the results of such an evaluation are presented for four collections. The results of retrieval experiments comparing a sequential search, a cluster-based search, and a search of the clustered collection in which individual documents are scored against the query are also presented.

These results indicate that while the absolute performance of a search on a particular collection is dependent on the pairwise similarity of the r ...

4 Combining classifiers in text categorization

Leah S. Larkey, W. Bruce Croft

August 1996 **Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(1.12 MB)   Additional Information: full citation, references, citings, index terms

5 An optimal algorithm for approximate nearest neighbor searching fixed dimensions

Sunil Arya, David M. Mount, Nathan S. Netanyahu, Ruth Silverman, Angela Y. Wu

November 1998 **Journal of the ACM (JACM)**, Volume 45 Issue 6

Publisher: ACM Press

Full text available: pdf(287.94 KB)   Additional Information: full citation, abstract, references, citings, index terms

Consider a set of S of n data points in real d-dimensional space, Rd, where distances are measured using any Minkowski metric. In nearest neighbor searching, we preprocess S into a data structure, so that given any query point $q \in Rd$, is the closest point of S to q can be reported quickly. Given any po ...

Keywords: approximation algorithms, box-decomposition trees, closet-point queries, nearest neighbor searching, post-office problem, priority search

6 Scatter/Gather: a cluster-based approach to browsing large document collections

Douglass R. Cutting, David R. Karger, Jan O. Pedersen, John W. Tukey

June 1992 **Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(1.08 MB)   Additional Information: full citation, abstract, references, citings, index terms

Document clustering has not been well received as an information retrieval tool. Objections to its use fall into two main categories: first, that clustering is too slow for large corpora (with running time often quadratic in the number of documents); and second, that clustering does not appreciably improve retrieval. We argue that these problems arise only when clustering is used in an attempt to improve conventional search techniques. However, looking at clustering as an informa ...

7 Searching in metric spaces

Edgar Chávez, Gonzalo Navarro, Ricardo Baeza-Yates, José Luis Marroquín

September 2001 **ACM Computing Surveys (CSUR)**, Volume 33 Issue 3

Publisher: ACM Press

Full text available: pdf(916.04 KB)   Additional Information: full citation, abstract, references, citings, index terms

The problem of searching the elements of a set that are close to a given query element under some similarity criterion has a vast number of applications in many branches of computer science, from pattern recognition to textual and multimedia information retrieval. We are interested in the rather general case where the similarity criterion defines a metric space, instead of the more restricted case of a vector space. Many solutions have been proposed in different areas, in many cases without cros ...

Keywords: Curse of dimensionality, nearest neighbors, similarity searching, vector spaces

8 Theory 1: Better than the real thing?: iterative pseudo-query processing using cluster-

based language models

Oren Kurland, Lillian Lee, Carmel Domshlak

August 2005 **Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR '05**

Publisher: ACM Press

Full text available: pdf(235.63 KB)   Additional Information: full citation, abstract, references, index terms

We present a novel approach to pseudo-feedback-based ad hoc retrieval that uses language models induced from both documents and clusters. First, we treat the pseudo-feedback documents produced in response to the original query as a set of *pseudo-query* that themselves can serve as input to the retrieval process. Observing that the documents returned in response to the pseudo-query can then act as pseudo-query for subsequent rounds, we arrive at a formulation of pseudo-query-based retrieval ...

Keywords: aspect recall, cluster-based language models, clustering, language modeling, pseudo-feedback, pseudo-queries, query drift, rendition

9   Dimensionality reduction: On scaling latent semantic indexing for large peer-to-peer systems

Chunqiang Tang, Sandhya Dwarkadas, Zhichen Xu

July 2004   **Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR '04**

Publisher: ACM Press

Full text available: pdf(208.57 KB)   Additional Information: full citation, abstract, references, index terms

The exponential growth of data demands scalable infrastructures capable of indexing and searching rich content such as text, music, and images. A promising direction is to combine information re-trieval with peer-to-peer technology for scalability, fault-tolerance, and low administration cost. One pioneering work along this di-rection is pSearch [32, 33]. pSearch places documents onto a peer-to- peer overlay network according to semantic vectors produced using Latent Semantic Indexing (LSI). The ...

Keywords: dimensionality reduction, latent semantic indexing, peer-to-peer IR

10   Search 2: Searching with numbers

Rakesh Agrawal, Ramakrishnan Srikant

May 2002   **Proceedings of the 11th international conference on World Wide Web**

Publisher: ACM Press

Full text available: pdf(238.60 KB)   Additional Information: full citation, abstract, references, citings, index terms

A large fraction of the useful web comprises of specification documents that largely consist of hattribute name, numeric valuei pairs embedded in text. Examples include product information, classified advertisements, resumes, etc. The approach taken in the past to search these documents by first establishing correspondences between values and their names has achieved limited success because of the difficulty of extracting this information from free text. We propose a new approach that does not r ...

11   Real-time shading

Marc Olano, Kurt Akeley, John C. Hart, Wolfgang Heidrich, Michael McCool, Jason L. Mitchell, Randi Rost

August 2004 **Proceedings of the conference on SIGGRAPH 2004 course notes GRAPH '04**

Publisher: ACM Press

Full text available: pdf(7.39 MB)   Additional Information: full citation, abstract

Real-time procedural shading was once seen as a distant dream. When the first version of this course was offered four years ago, real-time shading was possible, but only with one-of-a-kind hardware or by combining the effects of tens to hundreds of rendering passes. Today, almost every new computer comes with graphics hardware capable of interactively executing shaders of thousands to tens of thousands of instructions. This course has been

redesigned to address today's real-time shading capabili ...

12 Web page classification based on k-nearest neighbor approach

Oh-Woog Kwon, Jong-Hyeok Lee

November 2000 **Proceedings of the fifth international workshop on on Information retrieval with Asian languages**

Publisher: ACM Press

Full text available: pdf(653.68 KB)   Additional Information: full citation, abstract, references

Automatic categorization is the only viable method to deal with the scaling problem of the World Wide Web. In this paper, we propose a Web page classifier based on an adaptation of k-Nearest Neighbor (k-NN) approach. To improve the performance of k-NN approach, we supplement k-NN approach with a feature selection method and a term-weighting scheme using markup tags, and reform document-document similarity measure used in vector space model. In our experiments on a Korean commercial Web direct ...

Keywords: Web page classification, feature selection, k-nearest neighbor approach, similarity measure, term weighting scheme, text categorization

13 An adaptive *k*-nearest neighbor text categorization strategy

Li Baoli, Lu Qin, Yu Shiwen

December 2004 **ACM Transactions on Asian Language Information Processing (TALIP)**, Volume 3 Issue 4

Publisher: ACM Press

Full text available: pdf(285.94 KB)   Additional Information: full citation, abstract, references, index terms

k is the most important parameter in a text categorization system based on the k-nearest neighbor algorithm (kNN). To classify a new document, the k-nearest documents in the training set are determined first. The prediction of categories for this document can then be made according to the category distribution among the k nearest neighbors. Generally speaking, the class distribution in a training set is not even; some classes may have more samples than others. ...

Keywords: k-nearest neighbor algorithm, machine learning, text categorization, text classification

14 Language models: Corpus structure, language models, and ad hoc information retrieval

Oren Kurland, Lillian Lee

July 2004 **Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR '04**

Publisher: ACM Press

Full text available: pdf(214.33 KB)   Additional Information: full citation, abstract, references, index terms

Most previous work on the recently developed *language-modeling* approach to information retrieval focuses on document-specific characteristics, and therefore does not take into account the structure of the surrounding corpus. We propose a novel algorithmic framework in which information provided by document-based language models is enhanced by the incorporation of information drawn from *clusters* of similar documents. Using this framework, we develop a suite of new algorithms. Even t ...

Keywords: aspect models, cluster-based language models, clustering, interpolation model, language modeling, smoothing

15 Theory of keyblock-based image retrieval

April 2002 **ACM Transactions on Information Systems (TOIS)**, Volume 20 Issue 2

Publisher: ACM Press

Full text available: pdf(2.14 MB)   Additional Information: full citation, abstract, references, citings, index terms, review

The success of text-based retrieval motivates us to investigate analogous techniques which can support the querying and browsing of image data. However, images differ significantly from text both syntactically and semantically in their mode of representing and expressing information. Thus, the generalization of information retrieval from the text domain to the image domain is non-trivial. This paper presents a framework for information retrieval in the image domain which supports content-based q ...

Keywords: clustering, codebook, content-based image retrieval, keyblock

16 Cluster-based text categorization: a comparison of category search strategies
Makoto Iwayama, Takenobu Tokunaga
July 1995 **Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval**
Publisher: ACM Press
Full text available: pdf(721.57 KB)   Additional Information: full citation, references, citings, index terms

17 Evaluation of an inference network-based retrieval model
Howard Turtle, W. Bruce Croft
July 1991 **ACM Transactions on Information Systems (TOIS)**, Volume 9 Issue 3
Publisher: ACM Press
Full text available: pdf(2.40 MB)   Additional Information: full citation, references, citings, index terms, review

Keywords: document retrieval, inference networks, network retrieval models

18 An evaluation of phrasal and clustered representations on a text categorization task
David D. Lewis
June 1992 **Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval**
Publisher: ACM Press
Full text available: pdf(1.22 MB)   Additional Information: full citation, abstract, references, citings, index terms

Syntactic phrase indexing and term clustering have been widely explored as text representation techniques for text retrieval. In this paper we study the properties of phrasal and clustered indexing languages on a text categorization task, enabling us to study their properties in isolation from query interpretation issues. We show that optimal effectiveness occurs when using only a small proportion of the indexing terms available, and that effectiveness peaks at a higher feature set size and ...

19 GPGPU: general purpose computation on graphics hardware
David Luebke, Mark Harris, Jens Krüger, Tim Purcell, Naga Govindaraju, Ian Buck, Cliff Woolley, Aaron Lefohn
August 2004 **Proceedings of the conference on SIGGRAPH 2004 course notes GRAPH '04**
Publisher: ACM Press
Full text available: pdf(63.03 MB)   Additional Information: full citation, abstract

The graphics processor (GPU) on today's commodity video cards has evolved into an extremely powerful and flexible processor. The latest graphics architectures provide tremendous memory bandwidth and computational horsepower, with fully programmable vertex and pixel processing units that support vector operations up to full IEEE floating point precision. High level languages have emerged for graphics hardware, making this computational power accessible. Architecturally, GPUs are highly parallel s ...

20
Point-based computer graphics

Marc Alexa, Markus Gross, Mark Pauly, Hanspeter Pfister, Marc Stamminger, Matthias Zwicker

August 2004 **Proceedings of the conference on SIGGRAPH 2004 course notes GRAPH '04**

Publisher: ACM Press

Full text available: pdf(8.94 MB)   Additional Information: full citation, abstract

This course introduces points as a powerful and versatile graphics primitive. Speakers present their latest concepts for the acquisition, representation, modeling, processing, and rendering of point sampled geometry along with applications and research directions. We describe algorithms and discuss current problems and limitations, covering important aspects of point based graphics.

Results 1 - 20 of 200   Result page: **1** 2 3 4 5 6 7 8 9 10   next

Useful downloads: Adobe Acrobat   QuickTime   Windows Media Player   Real Player

P⊛RTAL
USPTO

THE ACM DIGITAL LIBRARY

⬤ Feedback Report a problem Satisfaction survey

Terms used
nearest neighbor cluster document precision recall

Found **197** of **169,166**

Sort results by   [relevance ▽]   **❧ Save results to a Binder**   Try an Advanced Search
Try this search in The ACM Guide

Display results   [expanded form ▽]   **⸮ Search Tips**
☐ Open results in a new window

Results 1 - 20 of 197   Result page: **1** 2 3 4 5 6 7 8 9 10   next

Relevance scale ☐☐▭▬■

1   IR-7 (information retrieval): natural language processing for IR: Stemming and
◈ lemmatization in the clustering of finnish text documents   ■
Tuomo Korenius, Jorma Laurikkala, Kalervo Järvelin, Martti Juhola
November 2004 **Proceedings of the thirteenth ACM international conference on
Information and knowledge management CIKM '04**
Publisher: ACM Press
Full text available: 📄 pdf(238.51 KB)   Additional Information: full citation, abstract, references, index terms

Stemming and lemmatization were compared in the clustering of Finnish text documents.
Since Finnish is a highly inflectional and agglutinative language, we hypothesized that
lemmatization, involving splitting of the compound words, would be more appropriate
normalization approach than the straightforward stemming. The relevance of the
documents were evaluated with a four-point relevance assessment scale, which was
collapsed into binary one by considering all the relevant and only the highly r ...

Keywords: clustering, lemmatization, normalization, stemming

2   Shape-based retrieval and analysis of 3D models   ■
◈ Thomas Funkhouser, Michael Kazhdan
August 2004 **Proceedings of the conference on SIGGRAPH 2004 course notes GRAPH
'04**
Publisher: ACM Press ,
Full text available: 📄 pdf(12.56 MB)   Additional Information: full citation, abstract

Large repositories of 3D data are rapidly becoming available in several fields, including
mechanical CAD, molecular biology, and computer graphics. As the number of 3D models
grows, there is an increasing need for computer algorithms to help people find the
interesting ones and discover relationships between them. Unfortunately, traditional text-
based search techniques are not always effective for 3D models, especially when queries
are geometric in nature (e.g., find me objects that fit into thi ...

3   The cluster hypothesis revisited   ■
◈ Ellen M. Voorhees
June 1985 **Proceedings of the 8th annual international ACM SIGIR conference on
Research and development in information retrieval**
Publisher: ACM Press
Full text available: 📄 pdf(559.94 KB)   Additional Information: full citation, abstract, references, citings

A new means of evaluating the cluster hypothesis is introduced and the results of such an
evaluation are presented for four collections. The results of retrieval experiments
comparing a sequential search, a cluster-based search, and a search of the clustered
collection in which individual documents are scored against the query are also presented.

These results indicate that while the absolute performance of a search on a particular collection is dependent on the pairwise similarity of the r ...

4 Combining classifiers in text categorization

Leah S. Larkey, W. Bruce Croft

August 1996 **Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(1.12 MB)   Additional Information: <u>full citation</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

5 An optimal algorithm for approximate nearest neighbor searching fixed dimensions

Sunil Arya, David M. Mount, Nathan S. Netanyahu, Ruth Silverman, Angela Y. Wu

November 1998 **Journal of the ACM (JACM)**, Volume 45 Issue 6

Publisher: ACM Press

Full text available: pdf(287.94 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

Consider a set of S of n data points in real d-dimensional space, Rd, where distances are measured using any Minkowski metric. In nearest neighbor searching, we preprocess S into a data structure, so that given any query point $q \in Rd$, is the closest point of S to q can be reported quickly. Given any po ...

Keywords: approximation algorithms, box-decomposition trees, closet-point queries, nearest neighbor searching, post-office problem, priority search

6 Scatter/Gather: a cluster-based approach to browsing large document collections

Douglass R. Cutting, David R. Karger, Jan O. Pedersen, John W. Tukey

June 1992 **Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(1.08 MB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

Document clustering has not been well received as an information retrieval tool. Objections to its use fall into two main categories: first, that clustering is too slow for large corpora (with running time often quadratic in the number of documents); and second, that clustering does not appreciably improve retrieval. We argue that these problems arise only when clustering is used in an attempt to improve conventional search techniques. However, looking at clustering as an informa ...

7 Searching in metric spaces

Edgar Chávez, Gonzalo Navarro, Ricardo Baeza-Yates, José Luis Marroquín

September 2001 **ACM Computing Surveys (CSUR)**, Volume 33 Issue 3

Publisher: ACM Press

Full text available: pdf(916.04 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

The problem of searching the elements of a set that are close to a given query element under some similarity criterion has a vast number of applications in many branches of computer science, from pattern recognition to textual and multimedia information retrieval. We are interested in the rather general case where the similarity criterion defines a metric space, instead of the more restricted case of a vector space. Many solutions have been proposed in different areas, in many cases without cros ...

Keywords: Curse of dimensionality, nearest neighbors, similarity searching, vector spaces

8 Theory 1: Better than the real thing?: iterative pseudo-query processing using cluster-

based language models

Oren Kurland, Lillian Lee, Carmel Domshlak

August 2005 **Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR '05**

Publisher: ACM Press

Full text available: pdf(235.63 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>index terms</u>

We present a novel approach to pseudo-feedback-based ad hoc retrieval that uses language models induced from both documents and clusters. First, we treat the pseudo-feedback documents produced in response to the original query as a set of *pseudo-query* that themselves can serve as input to the retrieval process. Observing that the documents returned in response to the pseudo-query can then act as pseudo-query for subsequent rounds, we arrive at a formulation of pseudo-query-based retrieval ...

Keywords: aspect recall, cluster-based language models, clustering, language modeling, pseudo-feedback, pseudo-queries, query drift, rendition

---

9 <u>Dimensionality reduction: On scaling latent semantic indexing for large peer-to-peer systems</u>

Chunqiang Tang, Sandhya Dwarkadas, Zhichen Xu

July 2004 **Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR '04**

Publisher: ACM Press

Full text available: pdf(208.57 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>index terms</u>

The exponential growth of data demands scalable infrastructures capable of indexing and searching rich content such as text, music, and images. A promising direction is to combine information re-trieval with peer-to-peer technology for scalability, fault-tolerance, and low administration cost. One pioneering work along this di-rection is pSearch [32, 33]. pSearch places documents onto a peer-to- peer overlay network according to semantic vectors produced using Latent Semantic Indexing (LSI). The ...

Keywords: dimensionality reduction, latent semantic indexing, peer-to-peer IR

---

10 <u>Search 2: Searching with numbers</u>

Rakesh Agrawal, Ramakrishnan Srikant

May 2002 **Proceedings of the 11th international conference on World Wide Web**

Publisher: ACM Press

Full text available: pdf(238.60 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

A large fraction of the useful web comprises of specification documents that largely consist of hattribute name, numeric valuei pairs embedded in text. Examples include product information, classified advertisements, resumes, etc. The approach taken in the past to search these documents by first establishing correspondences between values and their names has achieved limited success because of the difficulty of extracting this information from free text. We propose a new approach that does not r ...

---

11 <u>Real-time shading</u>

Marc Olano, Kurt Akeley, John C. Hart, Wolfgang Heidrich, Michael McCool, Jason L. Mitchell, Randi Rost

August 2004 **Proceedings of the conference on SIGGRAPH 2004 course notes GRAPH '04**

Publisher: ACM Press

Full text available: pdf(7.39 MB)   Additional Information: <u>full citation</u>, <u>abstract</u>

Real-time procedural shading was once seen as a distant dream. When the first version of this course was offered four years ago, real-time shading was possible, but only with one-of-a-kind hardware or by combining the effects of tens to hundreds of rendering passes. Today, almost every new computer comes with graphics hardware capable of interactively executing shaders of thousands to tens of thousands of instructions. This course has been

redesigned to address today's real-time shading capabili ...

12 Web page classification based on k-nearest neighbor approach

Oh-Woog Kwon, Jong-Hyeok Lee

November 2000 **Proceedings of the fifth international workshop on on Information retrieval with Asian languages**

Publisher: ACM Press

Full text available: pdf(653.68 KB)   Additional Information: full citation, abstract, references

Automatic categorization is the only viable method to deal with the scaling problem of the World Wide Web. In this paper, we propose a Web page classifier based on an adaptation of k-Nearest Neighbor (k-NN) approach. To improve the performance of k-NN approach, we supplement k-NN approach with a feature selection method and a term-weighting scheme using markup tags, and reform document-document similarity measure used in vector space model. In our experiments on a Korean commercial Web direct ...

Keywords: Web page classification, feature selection, k-nearest neighbor approach, similarity measure, term weighting scheme, text categorization

13 An adaptive *k*-nearest neighbor text categorization strategy

Li Baoli, Lu Qin, Yu Shiwen

December 2004 **ACM Transactions on Asian Language Information Processing (TALIP)**, Volume 3 Issue 4

Publisher: ACM Press

Full text available: pdf(285.94 KB)   Additional Information: full citation, abstract, references, index terms

k is the most important parameter in a text categorization system based on the k-nearest neighbor algorithm (kNN). To classify a new document, the k-nearest documents in the training set are determined first. The prediction of categories for this document can then be made according to the category distribution among the k nearest neighbors. Generally speaking, the class distribution in a training set is not even; some classes may have more samples than others. ...

Keywords: k-nearest neighbor algorithm, machine learning, text categorization, text classification

14 Language models: Corpus structure, language models, and ad hoc information retrieval

Oren Kurland, Lillian Lee

July 2004 **Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR '04**

Publisher: ACM Press

Full text available: pdf(214.33 KB)   Additional Information: full citation, abstract, references, index terms

Most previous work on the recently developed *language-modeling* approach to information retrieval focuses on document-specific characteristics, and therefore does not take into account the structure of the surrounding corpus. We propose a novel algorithmic framework in which information provided by document-based language models is enhanced by the incorporation of information drawn from *clusters* of similar documents. Using this framework, we develop a suite of new algorithms. Even t ...

Keywords: aspect models, cluster-based language models, clustering, interpolation model, language modeling, smoothing

15 Theory of keyblock-based image retrieval

April 2002 **ACM Transactions on Information Systems (TOIS)**, Volume 20 Issue 2

Publisher: ACM Press

Full text available: pdf(2.14 MB)   Additional Information: full citation, abstract, references, citings, index terms, review

The success of text-based retrieval motivates us to investigate analogous techniques which can support the querying and browsing of image data. However, images differ significantly from text both syntactically and semantically in their mode of representing and expressing information. Thus, the generalization of information retrieval from the text domain to the image domain is non-trivial. This paper presents a framework for information retrieval in the image domain which supports content-based q ...

Keywords: clustering, codebook, content-based image retrieval, keyblock

16 <u>Evaluation of an inference network-based retrieval model</u>

Howard Turtle, W. Bruce Croft

July 1991 **ACM Transactions on Information Systems (TOIS)**, Volume 9 Issue 3

Publisher: ACM Press

Full text available: pdf(2.40 MB)   Additional Information: <u>full citation</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>, <u>review</u>

Keywords: document retrieval, inference networks, network retrieval models

17 <u>Cluster-based text categorization: a comparison of category search strategies</u>

Makoto Iwayama, Takenobu Tokunaga

July 1995 **Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(721.57 KB)   Additional Information: <u>full citation</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

18 <u>An evaluation of phrasal and clustered representations on a text categorization task</u>

David D. Lewis

June 1992 **Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(1.22 MB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

Syntactic phrase indexing and term clustering have been widely explored as text representation techniques for text retrieval. In this paper we study the properties of phrasal and clustered indexing languages on a text categorization task, enabling us to study their properties in isolation from query interpretation issues. We show that optimal effectiveness occurs when using only a small proportion of the indexing terms available, and that effectiveness peaks at a higher feature set size and ...

19 <u>GPGPU: general purpose computation on graphics hardware</u>

David Luebke, Mark Harris, Jens Krüger, Tim Purcell, Naga Govindaraju, Ian Buck, Cliff Woolley, Aaron Lefohn

August 2004 **Proceedings of the conference on SIGGRAPH 2004 course notes GRAPH '04**

Publisher: ACM Press

Full text available: pdf(63.03 MB)   Additional Information: <u>full citation</u>, <u>abstract</u>

The graphics processor (GPU) on today's commodity video cards has evolved into an extremely powerful and flexible processor. The latest graphics architectures provide tremendous memory bandwidth and computational horsepower, with fully programmable vertex and pixel processing units that support vector operations up to full IEEE floating point precision. High level languages have emerged for graphics hardware, making this computational power accessible. Architecturally, GPUs are highly parallel s ...

20

<u>Point-based computer graphics</u>

Marc Alexa, Markus Gross, Mark Pauly, Hanspeter Pfister, Marc Stamminger, Matthias Zwicker

August 2004 **Proceedings of the conference on SIGGRAPH 2004 course notes GRAPH '04**

Publisher: ACM Press

Full text available: pdf(8.94 MB)   Additional Information: full citation, abstract

This course introduces points as a powerful and versatile graphics primitive. Speakers present their latest concepts for the acquisition, representation, modeling, processing, and rendering of point sampled geometry along with applications and research directions. We describe algorithms and discuss current problems and limitations, covering important aspects of point based graphics.

Results 1 - 20 of 197   Result page: **1** 2 3 4 5 6 7 8 9 10 next

Useful downloads: Adobe Acrobat   QuickTime   Windows Media Player   Real Player

Welcome United States Patent and Trademark Office

Search Results

BROWSE   SEARCH   IEEE XPLORE GUIDE   SUPPORT

Results for "((nearest and neighbor and cluster and document and precision and recall)<in>metadata)"   ✉ e-mail 🖨 printer friendly
Your search matched **0** documents.
A maximum of **100** results are displayed, **25** to a page, sorted by **Relevance** in **Descending** order.

» Search Options

View Session History

New Search

Modify Search

((nearest and neighbor and cluster and document and precision and recall)<in>metad »

☐ Check to search only within this results set

Display Format: ◉ Citation   ○ Citation & Abstract

» Key

IEEE JNL	IEEE Journal or Magazine
IEE JNL	IEE Journal or Magazine
IEEE CNF	IEEE Conference Proceeding
IEE CNF	IEE Conference Proceeding
IEEE STD	IEEE Standard

No results were found.

Please edit your search criteria and try again. Refer to the Help pages if you need assistance revising your search.

Indexed by
📕 Inspec®

Welcome United States Patent and Trademark Office

■▣▪Search Results

BROWSE   SEARCH   IEEE XPLORE GUIDE   SUPPORT

Results for "((nearest and neighbor and cluster and document and (precision or recall))<in>metadata)"   ☑ e-mail 🖶 printer friendly
Your search matched **0** documents.
A maximum of **100** results are displayed, **25** to a page, sorted by **Relevance** in **Descending** order.

» Search Options

View Session History

New Search

» Key

IEEE JNL	IEEE Journal or Magazine
IEE JNL	IEE Journal or Magazine
IEEE CNF	IEEE Conference Proceeding
IEE CNF	IEE Conference Proceeding
IEEE STD	IEEE Standard

Modify Search

((nearest and neighbor and cluster and document and (precision or recall))<in>metad ⏩

☐ Check to search only within this results set

Display Format: ◉ Citation ○ Citation & Abstract

No results were found.

Please edit your search criteria and try again. Refer to the Help pages if you need assistance revising your search.

Help   Contact Us   Privacy & Security   IEEE.org

Indexed by
繭 Inspec˙

IEEE *Xplore* ®
RELEASE 2.1

⌐□⌐Search Results

BROWSE   SEARCH   IEEE XPLORE GUIDE   SUPPORT

Results for "((nearest and neighbor and cluster and document)<in>metadata)"
Your search matched **13** of **1297674** documents.
A maximum of **100** results are displayed, **25** to a page, sorted by **Relevance** in **Descending** order.

✉ e-mail 🖨 printer friendly

» **Search Options**

View Session History

New Search

» **Key**

IEEE JNL	IEEE Journal or Magazine
IEE JNL	IEE Journal or Magazine
IEEE CNF	IEEE Conference Proceeding
IEE CNF	IEE Conference Proceeding
IEEE STD	IEEE Standard

Modify Search

((nearest and neighbor and cluster and document)<in>metadata) ⟩⟩

☐ Check to search only within this results set

Display Format: ◉ Citation   ○ Citation & Abstract

Select   Article Information

☐ 1. **Using topic keyword clusters for automatic document clustering**
Hsi-Cheng Chang; Chiun-Chieh Hsu;
Information Technology and Applications, 2005. ICITA 2005. Third International Conference on
Volume 1, 4-7 July 2005 Page(s):419 - 424 vol.1
Digital Object Identifier 10.1109/ICITA.2005.303

AbstractPlus | Full Text: PDF(224 KB)   **IEEE CNF**

☐ 2. **Skew detection of document images by focused nearest-neighbor clustering**
Xiaoyi Jiang; Bunke, H.; Widmer-Kljajo, D.;
Document Analysis and Recognition, 1999. ICDAR '99. Proceedings of the Fifth International
Conference on
20-22 Sept. 1999 Page(s):629 - 632
Digital Object Identifier 10.1109/ICDAR.1999.791866

AbstractPlus | Full Text: PDF(136 KB)   **IEEE CNF**

☐ 3. **A Web document clustering algorithm based on concept of neighbor**
Jiang-Chun Song; Jun-Yi Shen;
Machine Learning and Cybernetics, 2003 International Conference on
Volume 1, 2-5 Nov. 2003 Page(s):46 - 50 Vol.1

AbstractPlus | Full Text: PDF(507 KB)   **IEEE CNF**

☐ 4. **Classification and Clustering of Information Objects Based on Fuzzy Neighborhood
System**
Miyamoto, S.; Yasunori Endo; Hayakawa, S.; Kataoka, E.;
Systems, Man and Cybernetics, 2005 IEEE International Conference on
Volume 4, 10-12 Oct. 2005 Page(s):3210 - 3215

AbstractPlus | Full Text: PDF(136 KB)   **IEEE CNF**

☐ 5. **Improved document skew detection based on text line connected-component clustering**
Liolios, N.; Fakotakis, N.; Kokkinakis, G.;
Image Processing, 2001. Proceedings. 2001 International Conference on
Volume 1, 7-10 Oct. 2001 Page(s):1098 - 1101 vol.1
Digital Object Identifier 10.1109/ICIP.2001.959241

AbstractPlus | Full Text: PDF(248 KB)   **IEEE CNF**

☐ 6. **The document spectrum for page layout analysis**
O'Gorman, L.;
Pattern Analysis and Machine Intelligence, IEEE Transactions on
Volume 15, Issue 11, Nov. 1993 Page(s):1162 - 1173
Digital Object Identifier 10.1109/34.244677

AbstractPlus | Full Text: PDF(1108 KB)   **IEEE JNL**

7. **Multilingual text categorisation for global knowledge discovery using fuzzy techniques**

Chau, R.; Chung-Hsing Yeh;
Artificial Intelligence Systems, 2002. (ICAIS 2002). 2002 IEEE International Conference on
5-10 Sept. 2002 Page(s):82 - 86
Digital Object Identifier 10.1109/ICAIS.2002.1048058

AbstractPlus | Full Text: PDF(313 KB)   IEEE CNF

8. **Integrating knowledge components for writer identification in a digital archive of historical music scores**
Bruder, I.; Ignatova, T.; Milewski, L.;
Digital Libraries, 2004. Proceedings of the 2004 Joint ACM/IEEE Conference on
7-11 June 2004 Page(s):397
Digital Object Identifier 10.1109/JCDL.2004.1336176

AbstractPlus | Full Text: PDF(210 KB)   IEEE CNF

9. **Color map image segmentation using optimized nearest neighbor classifiers**
Yan, H.;
Document Analysis and Recognition, 1993., Proceedings of the Second International Conference on
20-22 Oct. 1993 Page(s):111 - 114
Digital Object Identifier 10.1109/ICDAR.1993.395770

AbstractPlus | Full Text: PDF(336 KB)   IEEE CNF

10. **Designing efficient distributed neural classifiers: application to handwritten digit recognition**
Ribert, A.; Lecourtier, Y.; Ennaji, A.;
Document Analysis and Recognition, 1999. ICDAR '99. Proceedings of the Fifth International Conference on
20-22 Sept. 1999 Page(s):265 - 268
Digital Object Identifier 10.1109/ICDAR.1999.791775

AbstractPlus | Full Text: PDF(56 KB)   IEEE CNF

11. **Match between normalization schemes and feature sets for handwritten Chinese character recognition**
Qing Wang; Zheru Chi; Feng, D.D.; Rongchun Zhao;
Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on
10-13 Sept. 2001 Page(s):551 - 555
Digital Object Identifier 10.1109/ICDAR.2001.953849

AbstractPlus | Full Text: PDF(416 KB)   IEEE CNF

12. **Recognition of essential folding operations: a step for interpreting illustrated books of origami**
Kato, J.; Watanabe, T.; Nakayama, T.;
Document Analysis and Recognition, 1997., Proceedings of the Fourth International Conference on
Volume 1, 18-20 Aug. 1997 Page(s):81 - 85 vol.1
Digital Object Identifier 10.1109/ICDAR.1997.619818

AbstractPlus | Full Text: PDF(376 KB)   IEEE CNF

13. **Static and dynamic classifier fusion for character recognition**
Prevost, L.; Milgram, M.;
Document Analysis and Recognition, 1997., Proceedings of the Fourth International Conference on
Volume 2, 18-20 Aug. 1997 Page(s):499 - 506 vol.2
Digital Object Identifier 10.1109/ICDAR.1997.620549

AbstractPlus | Full Text: PDF(480 KB)   IEEE CNF